Language complexity of rotations and Sturmian sequences

François Blanchard and Petr Kůrka

Abstract

Given a rotation of the circle, we study the complexity of formal languages that are generated by the itineraries of interval covers. These languages are regular iff the rotation is rational. In the case of irrational rotations, our study reduces to that of the language complexity of the corresponding Sturmian sequences. We show that for a large class of irrationals, including e, all quadratic numbers and more generally all Hurwitz numbers, the corresponding languages can be recognized by a nondeterministic Turing machine in linear time (in other words, belongs to **NLIN**).

1 Introduction

Suppose we are given a topological dynamical system, which means a continuous self-map T of a compact metric space X; let $\{C_0, C_1, \ldots C_{p-1}\}$ be a finite cover of X by closed sets, indexed by the finite alphabet $\{0, 1, \ldots, p-1\}$. An itinerary is the sequence of sets of such a cover visited by the trajectory of a point under T — or rather the sequence of their indices. The set of all infinite itineraries for a given cover is a one–sided subshift, and the set of all finite itineraries is the associated language. Itineraries thus establish a link between dynamics and languages.

To obtain a reasonable theory, one must restrict oneself to the simplest possible closed covers; otherwise the complexity of the obtained languages would be a property of the covers rather than of the dynamics. When X is a symbolic space of the form $A^{\mathbb{Z}}$ endowed with the product of discrete topologies on each coordinate, clopen partitions, i.e., partitions into closed-open or clopen sets, seem appropriate [10, 11]. When the space is the unit interval or the unit circle, the suitable covers consist of closed intervals overlapping at most in their endpoints, as in the case of irrational rotations of the circle and the associated Sturmian sequences [2], or unimodal maps of the unit interval [13].

Given a class of languages $\mathfrak{L}$, we say that a dynamical system is of class $\mathfrak{L}$ if there exists a separating sequence of open covers, all of them with associated languages in $\mathfrak{L}$; the rather technical definition of a separating sequence of covers is given in Section 2. A complexity class of languages is usually defined as the set of languages accepted by some type of abstract machine. Here we consider only nondeterministic machines: one could of course use deterministic ones, but then the computation times would be much greater; in the best case the results would be less striking, and in the worst one we should not be able to prove anything. The basic classes are **REG** (regular languages), **CF** (context-free languages), **NLIN** (languages recognized by a nondeterministic multitape Turing machine in linear time), and **REC** (recursive languages).

A stronger way for a dynamical system to be related to a language class is the following: suppose that for (X, T) there exists a finite closed cover (C_i), $i \in A$ such that for any infinite sequence (i_n) of symbols of A, $\bigcap_{n=0}^{\infty} T^{-n} C_{i_n}$ contains at most one element, so that there is a map from the set of infinite itineraries to X. In this case the set of itineraries is called a *symbolic representation* or *extension* of (X, T), and if it is of class $\mathfrak{L}$ we say that (X, T) admits a symbolic extension of class $\mathfrak{L}$.

An elementary example is that of expansions to the base 2: let X be the unit circle, T be the multiplication by 2 mod 1; then obviously the closed cover $\{[0, 1/2], [1/2, 0]\}$ has the right property, and the set $\{0, 1\}^{\mathbb{N}}$ of binary expansions of numbers on the circle, endowed with the shift, forms a symbolic representation of (X, T).

Dynamists are usually trying to construct nice symbolic extensions of dynamical systems [1, 2, 13], but there are situations where they are not appropriate tools.

In this article we consider rotations of the circle by α, and we study the computational complexity of languages generated by their itineraries from the point of view of Turing machines, according to arithmetical properties of α.

The first result is a rather natural one: a rotation of the circle is of class **REG** if and only if it is rational (Proposition 2). This does not mean it has a regular symbolic representation: actually what looks to us the simplest symbolic representation of the circle endowed with the identity map, the Grand Sturmian subshift, that is the set of all Sturmian sequences, is of class **NLIN** but not of class **REG** (Proposition 8)!

For the irrational rotation by the angle α the interval cover $([0, 1-\alpha], [1-\alpha, 0])$ that generates the associated Sturmian sequence is canonical in some sense (Proposition 4). The language complexity of an irrational rotation can therefore be defined as that of the corresponding Sturmian subshift.

Language classes of rotations present a gap: no irrational Sturmian subshift is of class **CF** (Proposition 3). A large family of irrational Sturmian subshifts is of class **NLIN**: this includes all those defined by Hurwitz numbers, which are irrationals whose continued fraction expansions are polynomial sequences (Proposition 6). The Hurwitz numbers in turn include all quadratics and the number e. All transcendental numbers of the form $\sum_{k=0}^{\infty} q^{-2^k}$, with $q \geq 2$ an integer, also generate Sturmian subshifts of class **NLIN**. There are also recursive

Sturmian subshifts that are not of class **NLIN**, and Sturmian subshifts that are not of class **REC**. Everything we know about the complexity of Sturmian subshifts associated with irrationals is based on the properties of their continued fraction expansions.

It is interesting to compare this set of results with some of the properties of β-shifts given in [3]. The β-shift is of class **REG** when β belongs to a class of algebraic integers containing the Pisot numbers; on the other hand whenever β is rational but not an integer the β-shift is not of class **REG**.

We are deeply grateful to P. Arnoux for some fruitful discussions, to B. Weiss for pointing out Shallit's article [17], and to the referee for a multitude of useful suggestions.

2 Dynamical systems, subshifts and languages

We start with basic concepts from topological dynamics as introduced, e.g., in [6]. We denote by $\overline{E}$ the closure of a subset E of a topological space; a set that is both closed and open is called *clopen* for short.

A *dynamical system* (X, f) is a nonempty compact metric space X endowed with a continuous self-map $f : X \to X$. The n^{th} *iteration* $f^n : X \to X$ of f is defined inductively by $f^0(x) = x$, $f^{n+1}(x) = f(f^n(x))$. A point $x \in X$ is *periodic*, if $f^n(x) = x$ for some $n > 0$. The smallest n with this property is called the period of x. A point $x \in X$ is *ultimately periodic* if it is not periodic but $f^m(x)$ is periodic for some $m > 0$. A set $Y \subseteq X$ is *invariant* if $f(Y) \subseteq Y$. If Y is also closed it is a compact metric space, and the dynamical system (Y, f) is called a *subsystem* of (X, f). (We use the same symbol f for the restriction of f to Y.) The *orbit* $o(x) = \{f^n(x); n \geq 0\}$ of a point $x \in X$ is an invariant set, and its closure $(\overline{o(x)}, f)$ is a subsystem of (X, f). A dynamical system is called *minimal* if it contains no proper subsystems. If (X, f) and (Y, g) are dynamical systems, a *homomorphism* $H : (X, f) \to (Y, g)$ is a continuous map $H : X \to Y$ such that $Hf = gH$. A bijective homomorphism is called a *conjugacy* and a surjective homomorphism is called a *factor map*. If $H : (X, f) \to (Y, g)$ is a factor map, then we say that (Y, g) is a *factor* of (X, f), or that (X, f) is an *extension* of (Y, g).

Our main examples of dynamical systems are the rotations of the circle. Parametrize the circle by the semiopen interval $T_1 = [0, 1)$. The distance between two points is the length of the shorter arc between them:

$$d(x, y) = \min\{|x - y|,\ |1 + x - y|,\ |1 + y - x|\}.$$

Given a real number $\alpha \in T_1$, the rotation of the circle by α is the map defined on T_1 by $f_\alpha(x) = x + \alpha \bmod 1$ (here $z = y \bmod 1$ if $z \in T_1$ and $z - y$ is an integer). Rotations are continuous (in fact $d(f_\alpha(x), f_\alpha(y)) = d(x, y)$), so (T_1, f_α) is a dynamical system for every α. If $\alpha = p/q$ is a rational number with p, q coprime and $q > 0$, then every point $x \in T_1$ is periodic with period q. If α

is irrational, then (T_1, f_α) has no periodic points and it is minimal. Observe that for every pair of real numbers α, β, the map $f_\beta : (T_1, f_\alpha) \to (T_1, f_\alpha)$ is a conjugacy.

Other dynamical systems we are concerned with are symbolic systems or subshifts. They are closely related to languages, and before defining them we must introduce some language-theoretic definitions and notations. If A is a finite alphabet, $n \in \mathbb{N} = \{0, 1, 2, \ldots\}$, denote by A^n the set of words over A of length n, $A^* = \cup_{n \in \mathbb{N}} A^n$ the set of finite words over A, $A^{\mathbb{N}}$ the set of simply infinite words, and $\overline{A^*} = A^* \cup A^{\mathbb{N}}$. For $u \in \overline{A^*}$, denote by $|u|$ its length $(0 \leq |u| \leq \infty)$, and $|u|_a$ the number of occurrences of a letter $a \in A$ in u. Call λ the empty word. The $(i+1)^{st}$ letter of a word $u \in \overline{A^*}$ is denoted by u_i, so $u = u_0 u_1 \cdots$; when $i > |u|$ put $u_i = \lambda$. For $u \in A^*$, $v \in \overline{A^*}$ write $u \sqsubseteq v$ if u is a *subword* of v, i.e., if there exists $j \geq 0$ such that $u_i = v_{j+i}$ for all $i < |u|$; this is what is usually called a factor in language theory, but we are already using this term for homomorphic surjective images of dynamical systems. Denote by $u_{|i} = u_0 \cdots u_{i-1}$ the *initial subword* (or *prefix*) of u of length i. The *concatenation* of the words u and v is written uv; denote the n^{th} concatenation power of u by u^n, and the periodic sequence generated by its infinite repetition by $u^\infty \in A^{\mathbb{N}}$. We frequently use the binary alphabet $\mathbf{2} = \{0, 1\}$.

Define a metric d on the power space $A^{\mathbb{N}}$ by

$$d(u, v) = 2^{-n}, \quad \text{where} \quad n = \min\{i \in \mathbb{N}; u_i \neq v_i\}$$

(this metric can be extended by the same formula to $\overline{A^*}$, which becomes then the closure of A^*; this explains the notation). The space $A^{\mathbb{N}}$ is compact and homeomorphic to the Cantor middle third set. For $u \in A^*$ denote by

$$[u] = \{v \in A^{\mathbb{N}}; (\forall i < |u|)(u_i = v_i)\}$$

the *cylinder* of u. It is a clopen set. The space $A^{\mathbb{N}}$ is *zero-dimensional*: this means that if $U \subseteq A^{\mathbb{N}}$ is an open set and $u \in U$, then there exists $n \in \mathbb{N}$ such that $[u_{|n}] \subseteq U$. In general a compact metric space is zero-dimensional iff it is homeomorphic to a subspace of $A^{\mathbb{N}}$. A zero-dimensional dynamical system is one that is defined on a zero-dimensional space.

The shift map $\sigma : A^{\mathbb{N}} \to A^{\mathbb{N}}$ is defined by $\sigma(u)_i = u_{i+1}$; thus $\sigma(u_0 u_1 u_2 \cdots) = u_1 u_2 u_3 \cdots$. Then $d(\sigma(u), \sigma(v)) \leq 2d(u, v)$, so σ is continuous, and $(A^{\mathbb{N}}, \sigma)$ is a dynamical system; it is called a *full shift*.

A subsystem of a full shift is called a *subshift*. For instance the set $X = \{x \in \{0,1\}^{\mathbb{N}};\ \forall i \in \mathbb{N},\ x_i x_{i+1} \neq 11\} \subset \{0,1\}^{\mathbb{N}}$ is a subshift, because it is closed and shift-invariant. Here X is defined by a condition on words: x belongs to X if and only if the word 11 never occurs as a block of two consecutive coordinates. Actually this is a general feature of subshifts. Given a subshift $S \subseteq A^{\mathbb{N}}$, define its *associated language* as

$$\mathcal{L}(S) = \{u \in A^* \; ; (\exists v \in S)(u \sqsubseteq v)\}.$$

The language $\mathcal{L}(S)$ is *right central*, i.e., closed under subwords (if $u \in \mathcal{L}(S)$ and $v \sqsubseteq u$ then $v \in \mathcal{L}(S)$) and extendable to the right (if $u \in \mathcal{L}(S)$, then there exists $a \in A$ such that the concatenation ua is in $\mathcal{L}(S)$). The converse is true: any right central language is the associated language of a unique subshift (cf.[4]). Subshifts are examples of zero-dimensional dynamical systems

Finite closed covers are a major tool in this article, because they allow one to construct zero-dimensional extensions of dynamical systems. If X is a compact metric space, a *finite closed cover* of X is a family $\mathcal{V} = \{V_a \ ; a \in A\}$ of closed subsets of X indexed on the finite alphabet A, such that the union of all V_a is the full space X; these subsets are not necessarily disjoint. The *diameter* of $\mathcal{V}$ is $\mathrm{diam}(\mathcal{V}) = \max\{\mathrm{diam}(V_a); a \in A\}$; the diameter of a set $Y \subseteq X$ is $diam(Y) = \sup\{d(y,z); y, z \in Y\}$. A cover $\mathcal{V}$ is said to be *finer* than a cover $\mathcal{W}$ if any element of $\mathcal{V}$ is contained in some element of $\mathcal{W}$. We say that a sequence of finite closed covers $(\mathcal{V}_i)$, $i \in \mathbb{N}$ is *separating* if $\mathcal{V}_{i+1}$ is finer than $\mathcal{V}_i$ and $\lim_{i\to\infty} \mathrm{diam}(\mathcal{V}_i) = 0$.

A *clopen partition* of a zero-dimensional space is a finite closed cover consisting of disjoint clopen sets. The product space $A^{\mathbb{N}}$ has a natural clopen partition $\mathcal{V}_1 = \{[a] \ ; a \in A\}$; one often considers the separating sequence of clopen partitions $\mathcal{V}_n = \{[u]; u \in A^n\}$.

An *interval cover* of the circle is a finite cover consisting of at least two closed nondegenerate intervals intersecting at most in their endpoints. Among closed intervals we also include $[a,b] = [a,1) \cup [0,b]$ when $0 \le b < a < 1$. One can write an interval cover as $\mathcal{V} = \{V_a; a \in A\}$, where $A = \{0, 1, \ldots, n-1\}$, $n \ge 2$, $V_a = [c_a, c_{a+1}]$ for $a < n-1$, and $V_{n-1} = [c_{n-1}, c_0]$ for some sequence $0 \le c_0 < c_1 < \cdots < c_{n-1} < 1$.

Let (X, f) be a dynamical system and $\mathcal{V} = \{V_a; a \in A\}$ be a finite closed cover of X. For $u \in \overline{A^*}$ put

$$V_u = \{x \in X; (\forall i < |u|)(f^i(x) \in V_{u_i})\} = \bigcap_{i<|u|} f^{-i}(V_{u_i}).$$

Clearly every V_u is a closed set. The inclusion $x \in V_u$ means that u is an *itinerary* of x obtained from the sequence of iterates $f^i(x)$ by noting for each i a set of the cover to which $f^i(x)$ belongs. In general a given point has several itineraries because the sets of the cover overlap. When $\mathcal{V}$ is a clopen partition every point has a unique itinerary. Put

$$\mathcal{L}_{\mathcal{V}}(X,f) = \{u \in A^*; V_u \ne \emptyset\},\ S_{\mathcal{V}}(X,f) = \{u \in A^{\mathbb{N}}; V_u \ne \emptyset\};$$

Then $S_{\mathcal{V}}(X,f)$ is a subshift and $\mathcal{L}_{\mathcal{V}}(X,f)$ is its language. We say that a finite closed cover $\mathcal{V} = \{V_a; a \in A\}$ is a *generator* for (X,f) if for all $u \in A^{\mathbb{N}}$, V_u contains at most one point. Then the map $H : (S_{\mathcal{V}}(X,f), \sigma) \to (X,f)$ defined by $H(u) \in V_u$ is a factor map, and $(S_{\mathcal{V}}(X,f), \sigma)$ is called a *symbolic extension* of (X,f). In the Introduction we introduced the example of binary expansions. Another, more complex classical instance is any irrational rotation of the circle, for which any interval cover is a generator; in this case the set of itineraries is a proper subshift.

In particular when S is a subshift the canonical clopen partition $\mathcal{V}_1 = \{[a]; a \in A\}$ is a generator and $\mathcal{L}_{\mathcal{V}}(S, \sigma) = \mathcal{L}(S)$ is just the language of words occurring in elements of S.

Here is an example taken from [10]. A (one-sided) cellular automaton is a dynamical system $(A^{\mathbb{N}}, f)$ defined by $(f(u))_i = f_0(u_i, \ldots, u_{i+r})$, where $r \in \mathbb{N}$ and $f_0 : A^{r+1} \to A$ is a given *local rule.* Then $\mathcal{L}_{\mathcal{V}_1}(X, f)$ consists of all words $f^i(x),\ f^{i+1}(x), \ldots, f^{i+k}(x)$ $(x \in X)$ that occur as sequences of consecutive values of the coordinate 0 under the action of f. For example when $A = \mathbf{2}$ and f is given by $f(u)_i = u_i \cdot u_{i+1}$ (multiplication), then $\mathcal{L}_{\mathcal{V}_1}(X, f) = \{1^n 0^m; n, m \geq 0\}$. If f is given by $f(u)_i = u_i + u_{i+1} \bmod 2$, then $\mathcal{L}_{\mathcal{V}_1}(X, f) = \mathbf{2}^*$ is the full shift.

3 Language classes

A language over an alphabet A is any subset $L \subseteq A^*$. We distinguish several classes of languages defined by different kinds of computational devices (see Hopcroft & Ullman [8], Wagner & Wechsung [18] or Papadimitriou [15]).

A language is regular (**REG**) if it can be recognized by a (deterministic or nondeterministic) finite automaton. A very close equivalent definition, slightly handier in symbolic dynamics, uses finite graphs. A finite labeled graph over an alphabet A is a quintuple $G = (V, E, s, t, l)$, where V is a finite set of vertices, E is a finite set of edges, $s, t : E \to V$ are the source and target maps and $l : E \to A$ is the labeling function (cf. Lind & Marcus [12], p.64). A finite or infinite *path* in G is any sequence $e \in \overline{E^*}$ such that $s(e_i) = t(e_{i-1})$ for any $0 < i < |e|$. The label of a path $e \in \overline{E^*}$ is $l(e_0)l(e_1)\ldots \in \overline{A^*}$. A right central language L is *regular* iff it is the set of finite path labels for some finite labeled graph G; then we say that G recognizes L.

For dynamical purposes, two subclasses of the class of regular languages seem to be useful. We say that a (right central) regular language L is periodic (**PER**) if the corresponding subshift is countable. L is bounded periodic (**BPER**) (if there is a graph G recognizing L) if the corresponding subshift is finite (consists of a finite collection of periodic points). The corresponding purely language-theoretic definitions are not very complicated; finding them is a good exercise for the interested reader. A typical periodic right central language that is not bounded periodic is $L = \{a^* b^*\}$.

A language is context-free (**CF**) if it can be recognized by a nondeterministic push-down automaton. Higher complexity classes are defined as sets of languages accepted by nondeterministic Turing machines with possible constraints on the time or space. A nondeterministic Turing machine with n tapes is given by an input alphabet A_0, work alphabets $A_1, \ldots, A_n$, a finite set of states Q, an initial state $q_0 \in Q$, a set of final states $Q_1 \subseteq Q$, a set of accepting states

$Q_0 \subseteq Q_1$, and a transition function

$$\delta : Q \times (A_0 \cup \{\lambda\}) \times \cdots \times (A_n \cup \{\lambda\}) \to \mathcal{P}(Q \times A_1 \times \cdots \times A_n \times \{0,1\} \times \{-1,0,1\}^n);$$

here $\mathcal{P}(X) = \{Y; Y \subseteq X\}$ is the power set of X.

In this definition final states do not play the same rôle as in the classical definition of regular languages by way of automata: here they just mean that if the input tape is empty the computation is finished. Let us explain this in more details. A configuration of a Turing machine is a $(2n+2)$-tuple $(q, u^{(0)}, u^{(1)}, v^{(1)}, \cdots, u^{(n)}, v^{(n)})$ where $q \in Q$, $u^{(0)} \in A_0^*$, and $u^{(i)}, v^{(i)} \in A_i^*$. This means that the (zeroth) input tape contains the word $u^{(0)}$ with the head pointing to its first letter $u_0^{(0)}$ and the i^{th} work tape contains the concatenation $v^{(i)}u^{(i)}$ with the head pointing to the first letter of $u^{(i)}$. A configuration is initial if its state is q_0 and all work tapes are empty; it is final if its state belongs to Q_1; it is accepting if its state is in Q_0 and its input tape is empty ($u^{(0)} = \lambda$). A computation is a sequence of configurations starting at an initial configuration and that is compatible with the transition function δ. This means that the state q of the configuration and all the letters read on the tapes by the heads determine via δ one of the elements in the power set and the configuration is updated accordingly. The state is changed, the fields under heads on the work tapes are rewritten, the input head rests (0) or moves to the right (1) and the i^{th} head moves to the left(-1), rests (1) or moves to the right (1).

A computation is either infinite when it contains no final configuration, or finite when its last (and only its last) configuration is final. A word $u^{(0)} \in A_0^*$ is accepted (rejected) by the Turing machine if there exists a finite computation starting in the initial configuration with $u^{(0)}$ on the input tape, whose last configuration is accepting (not accepting).

A language is said to be recursive (**REC**) if it is the set of words accepted by some Turing machine that makes a finite computation on every input word. A language L is nondeterministic linear time (**NLIN**) if it is recursive for a Turing machine with the following property: there is a constant $c \geq 1$ such that for every $u \in L$ there exists an accepting computation with length at most $c|u|$ starting at an initial configuration with u on the input tape. Book & Greibach [5] proved that every nondeterministic linear time language is nondeterministic real time, which means linear time with constant $c = 1$. More generally to any proper complexity function $f(n)$ one associates the class $\mathbf{NTIME}(f(n))$ of languages accepted by nondeterministic multitape Turing machines in time $f(|u|)$ [15]. One has the following inclusions for right central languages:

$$\mathbf{BPER} \subset \mathbf{PER} \subset \mathbf{REG} \subset \mathbf{CF} \subset \mathbf{NLIN} \subset \mathbf{REC}.$$

Definition 1 *A family $\mathfrak{L}$ of languages is closed under factors if for any factor map $H : (S_1, \sigma) \to (S_2, \sigma)$ between subshifts, if $\mathcal{L}(S_1)$ belongs to $\mathfrak{L}$, then $\mathcal{L}(S_2)$ belongs to $\mathfrak{L}$. A family of languages $\mathfrak{L}$ is closed under concatenations if for* *every language $L \subseteq A^*$ of class $\mathfrak{L}$, and every $n > 0$, the language*

$L^n = \{(u_0 \cdots u_{n-1})(u_1 \cdots u_n) \ldots (u_{k-1} \cdots u_{k+n-2}) \in (A^n)^*; u_0 \cdots u_{k+n-2} \in L)\}$

is in $\mathfrak{L}$.

All classes of languages considered above are closed both under factors and concatenations; in the sequel, any abstract class of languages considered is supposed to have these properties.

Definition 2 *Let $\mathfrak{L}$ be a class of languages. Define a zero-dimensional system (X, f) to be of class $\mathfrak{L}$ if there exists a separating sequence of clopen partitions $\mathcal{V}_i$ of X such that for every i, $\mathcal{L}_{\mathcal{V}_i}(X, f)$ is in $\mathfrak{L}$. A dynamical system (T_1, f) isof class $\mathfrak{L}$ if there exists a separating sequence of interval covers $\mathcal{V}_i$ of X such that for every i, $\mathcal{L}_{\mathcal{V}_i}(X, f)$ is in $\mathfrak{L}$.*

Proposition 1 *Every dynamical system (T_1, f) of class $\mathfrak{L}$ is a factor of a zero-dimensional system of class $\mathfrak{L}$.*

Proof: This works for a general dynamical system (X, f) possessing a separating sequence of closed covers $\mathcal{V}_i = \{V_a; a \in A_i\}$. Since $\mathcal{V}_{i+1}$ is finer than $\mathcal{V}_i$, there exists a map $h_i : A_{i+1} \to A_i$ with $V_a \subseteq V_{h_i(a)}$ for every $a \in A_{i+1}$. We have factor maps $h_i^{I\!N} : (Y_{i+1}, \sigma) \to (Y_i, \sigma)$, where $Y_i = S_{\mathcal{V}_i}(X, f)$, and $h_i^{I\!N}(u)_i = h(u_i)$. Let $Y = \{y \in \prod_i Y_i;\ h_i^{I\!N}(y_{i+1}) = y_i\}$ be their inverse limit, define $g : Y \to Y$ by $g(y)_i = \sigma(y_i)$, and $H : Y \to X$ by $H(y) = \cap_i V_{y_{i0}}$. (Here $y_i \in Y_i$ is the i^{th} component of $y \in Y$ and $y_{i0} \in A_i$ is the zeroth component of y_i.) Then (Y, g) is a zero-dimensional system of class $\mathfrak{L}$, and $H : (Y, g) \to (X, f)$ is a factor map. $\Box$

4 Sturmian sequences and subshifts

Sturmian sequences were introduced by Morse and Hedlund [14]. Here are some facts about them that are relevant to our purpose. A sequence $u \in \mathbf{2}^{I\!N}$ is called *Sturmian* if in any two subwords v, w of u of the same length the numbers of occurrences of 0 differ at most by one: $||v|_0 - |w|_0| \leq 1$. Sturmian sequences can be described with the help of continued fractions. Let $I\!N_+ = \{1, 2, \ldots\}$ be the set of positive integers, denote by $I\!N_+^*$ the set of finite sequences, $I\!N_+^{I\!N}$ the set of infinite sequences and $\overline{I\!N_+^*} = I\!N_+^* \cup I\!N_+^{I\!N}$ the set of finite or infinite sequences of positive integers. For $a = (a_i)_{0 \leq i < n} \in \overline{I\!N_+^*}$ let $\langle a \rangle$ be the real number with continued fraction expansion

$$\langle a_0, a_1, \ldots \rangle = \cfrac{1}{a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}}$$

For the empty sequence $\lambda \in \overline{I\!N_+^*}$ one has $\langle\lambda\rangle = 0$; also $\langle 1\rangle = 1$, $\langle 1,1\rangle = \langle 2\rangle = 1/2$, $\langle 1^\infty\rangle = (\sqrt{5}-1)/2$. It is well known that $a \in I\!N_+^{I\!N}$ is periodic or ultimately periodic iff $\langle a\rangle$ is a zero of a quadratic polynomial with integer coefficients. With every $a \in \overline{I\!N_+^*}$ we associate a sequence of words $W_n \in \mathbf{2}^*$ by

$$W_{-1} = 1,\ W_0 = 0,\ W_1 = 0^{a_0-1}1,\ W_2 = W_1^{a_1}W_0, \dots\ W_{n+1} = W_n^{a_n}W_{n-1}.$$

For $|a| = \infty$, put

$$W(a) = \lim_{n\to\infty} W_n, \quad S_a = \overline{\{\sigma^i(W(a)); i \geq 0\}},$$

where σ is the shift defined in Section 2, and overlined symbols denote the closures of the corresponding sets. For $|a| = n$, put

$$W(a) = (W_n)^\infty, \quad S_a = \overline{\{\sigma^i(W_n^k W_{n-1}(W_n)^\infty);\ k \geq 0,\ i \geq 0\}}.$$

Thus $W(\lambda) = 0^\infty$, $S_\lambda = \{0^k10^\infty; k \geq 0\} \cup \{0^\infty\}$, $W(1) = 1^\infty$, $S_1 = \{1^k01^\infty; k \geq 0\} \cup \{1^\infty\}$, $W(2) = (01)^\infty$, $W(1,1) = (10)^\infty$ and

$$W(1^\infty) = 1\ 0\ 1\ 10\ 101\ 10110\ 10110101\ 1011010110110 \cdots$$

If a is finite, then S_a consists of periodic and ultimately periodic sequences. If $|a| = n$, $a_{n-1} = 1$, and $b = (a_0, \dots, a_{n-3}, a_{n-2}+1)$, then S_a and S_b contain the same periodic sequences but different preperiodic sequences. With this exception, $S_a \cap S_b = \emptyset$ for $a \neq b$. Note also that if $1/2 < \langle a\rangle \leq 1$, then $a_0 = 1$, and $1 - \langle a\rangle = \langle a_1+1, a_2, \dots\rangle$. It follows that $H : (S_a, \sigma) \to (S_{(a_1+1,a_2,\dots)}, \sigma)$ defined by $H(u)_i = 1 - u_i$ is a conjugacy.

A sequence is Sturmian iff it belongs to some S_a [2]. Every Sturmian sequence u has a well-defined frequency $\mu(u) = \lim_{n\to\infty} \mathrm{card}\{i < n; u_i = 1\}/n$, and for every $a \in \overline{I\!N_+^*}$, $\mu(S_a) = \langle a\rangle$. If a is infinite, and so $\langle a\rangle$ is irrational, then S_a is exactly the set of Sturmian sequences with frequency $\alpha = \langle a\rangle$ and

$$\begin{array}{lll} W(a)_i = 0 & \text{iff} & (\exists n \in I\!N)(0 < (i+1)\langle a\rangle - n < 1 - \langle a\rangle) \\ W(a)_i = 1 & \text{iff} & (\exists n \in I\!N)(1 - \langle a\rangle < (i+1)\langle a\rangle - n < 1). \end{array}$$

This just means that for $\alpha = \langle a\rangle$, $W(a)$ is the itinerary of the point 0 for the rotation (T_1, f_α) and its canonical interval cover $\mathcal{V}_\alpha = \{[0, 1-\alpha], [1-\alpha, 0]\}$. It follows that $S_{\mathcal{V}_\alpha}(T_1, f_\alpha) = S_a$. This geometric interpretation of Sturmian sequences is one of the main motivations for their study.

We call the set of all Sturmian sequences $S_\infty = \cup\{S_a; a \in \overline{I\!N_+^*}\}$ the *Grand Sturmian* subshift. One easily checks that $S_\infty \subseteq \mathbf{2}^{I\!N}$ is a closed σ-invariant set. The frequency map μ is continuous on S_∞, indeed it is a factor map from (S_∞, σ) to $([0,1], Id)$.

5 Rotations of the circle

A rotation of the circle is a dynamical system (T_1, f_α) where $f_\alpha(x) = x+\alpha \bmod 1$ (see Section 2). Now we investigate the language complexity of rotations as defined in Section 3. Call a rotation rational if α is rational. The next Proposition shows that for rational rotations there exists a separating sequence of interval covers $\mathcal{V}_i$ such that for every i, $\mathcal{L}_{\mathcal{V}_i}(T_1, f_\alpha)$ is regular, but there exists none such that $\mathcal{L}_{\mathcal{V}_i}(T_1, f_\alpha)$ is a periodic language; this proves the converse of Proposition 1 is not true.

Proposition 2 *Every rational rotation is of class* **REG** *but never of class* **PER**. *However, it is a factor of a zero-dimensional system of class* **BPER**.

Proof: Let $\alpha = p/q$, and m be a multiple of q. Put $\mathcal{V} = \{[\frac{i}{m}, \frac{i+1}{m}];\ \ 0 \leq i < m\}$. Then $\mathcal{L}_\mathcal{V}(S_1, f_\alpha)$ can be obtained as the language of a labelled graph $G = (V, E, s, t, l)$, where $V = \{0, \ldots, m-1\}$, $E = V \times \{0, 1\}$, $s(i, j) = i$, $t(i, j) = i + 1 \bmod m$, $l(r, s) = s + rpm/q \bmod m$; there must be two edges from each vertex because of the endpoints of the intervals (see Fig. 1 for the case $p = 1$, $m = q$). Thus (T_1, f_α) is of class **REG**. To show that it is not of class **PER**, consider any finite closed cover $\mathcal{V} = \{V_a; a \in A\}$. There exist $a \neq b$ and $x \in V_a \cap V_b$. Let $u \in A^{\mathbb{N}}$ be such that $x \in V_u$. Then any $v \in A^{\mathbb{N}}$ with $v_{iq} \in \{a, b\}$ and $v_j = u_j$ whenever j is not a multiple of q belongs to $S_\mathcal{V}(T_1, f_\alpha)$: so $S_\mathcal{V}(T_1, f_\alpha)$ is not countable. Finally $(T_1, f_{p/q})$ is a factor of a zero-dimensional dynamical system of class **BPER**: consider $(A^{\mathbb{N}}, g)$, where $A = \{0, \ldots, q-1\}$, $g(x)_0 = x_0 + 1 \bmod q$, and $g(x)_i = x_i$ for $i > 0$. As $H : \{u \in A^{\mathbb{N}}; u_0 = 0\} \to [0, 1/q]$ is a continuous surjective map it can be extended uniquely to a factor map $H : (A^{\mathbb{N}}, g) \to (T_1, f_\alpha)$. □

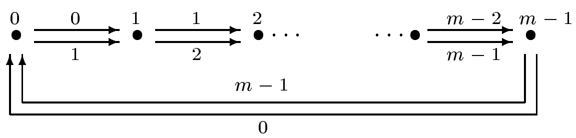

Fig. 1

A rational rotation $(T_1, f_{p/q})$ is obviously a factor of the subshift $S_\infty \times \{\sigma^i(\overline{1^{q-1}0}); 0 \leq i < q\}$.

Let us now deal with the much more interesting irrational case. First here is a kind of "lower bound" for the complexity of irrational rotations.

Proposition 3 *No irrational rotation of the circle is of class* **CF**.

Proof: We show that every subshift of class **CF** has a periodic point. Let $S \subseteq A^{\mathbb{N}}$ be a **CF** subshift. Since its language $\mathcal{L}(S)$ is infinite, by the Pumping Lemma ([8], p.125) there exists $w = uvxyz \in \mathcal{L}(S)$, such that $|v| + |y| \geq 1$, and

$uv^nxy^nz \in \mathcal{L}(S)$ for every $n \in I\!N$. If $|v| \geq 1$, then $v^\infty \in S$ is a periodic point. If $|y| \geq 1$, then $y^\infty \in S$ is a periodic point. Suppose that (T_1, f_α) is an irrational rotation and $\mathcal{V}$ an interval cover such that $\mathcal{L}_V(T_1, f_\alpha)$ is **CF**. Since every interval cover is a generator, (T_1, f_α) is a factor of $S_\mathcal{V}(T_1, f_\alpha)$, so it contains a periodic point, which is a contradiction. $\square$

Proposition 4 *Let $0 < \alpha < 1$ be irrational, and let $\mathcal{V}$ be any interval cover. Then $S_{\mathcal{V}_\alpha}(T_1, f_\alpha)$ is a factor of $S_\mathcal{V}(T_1, f_\alpha)$.*

Proof: Since $f_\beta : (T_1, f_\alpha) \to (T_1, f_\alpha)$ is a conjugacy for any β, we can suppose that $\mathcal{V}$ has 0 among its endpoints. Then the closed partition $\{V_a \cap f_\alpha^{-1}(V_b); a, b \in A\}$ is finer than $\mathcal{V}_\alpha$. Define a factor map $H : (S_\mathcal{V}(T_1, f_\alpha), \sigma) \to (S_{\mathcal{V}_\alpha}(T_1, f_\alpha), \sigma)$ by $H(u)_i = 0$ if $V_{u_i} \cap f_\alpha^{-1}(V_{u_{i+1}}) \subseteq [0, 1-\alpha]$, and $H(u)_i = 1$ if $V_{u_i} \cap f_\alpha^{-1}(V_{u_{i+1}}) \subseteq [1-\alpha, 0]$. $\square$

Theorem 1 *Let $\mathfrak{L}$ be a class of languages closed under factors and concatenations. Then (T_1, f_α) is of class $\mathfrak{L}$ iff $S_{\mathcal{V}_\alpha}(T_1, f_\alpha)$ is of class $\mathfrak{L}$.*

Proof: If (T_1, f_α) is of class $\mathfrak{L}$, then $S_{\mathcal{V}_\alpha}(T_1, f_\alpha)$ is of class $\mathfrak{L}$ by Proposition 4. Conversely suppose that $S_{\mathcal{V}_\alpha}(T_1, f_\alpha)$ is of class $\mathfrak{L}$. For $n > 0$ put $A_n = \mathcal{L}_\mathcal{V}(T_1, f_\alpha) \cap \mathbf{2}^n$. Then $\mathcal{V}_n = \{V_u; u \in A_n\}$ is an interval cover, and $\mathcal{L}_{\mathcal{V}_n}(T_1, f_\alpha) = \mathcal{L}^n_\mathcal{V}(T_1, f_\alpha)$ belongs to $\mathfrak{L}$, since $\mathfrak{L}$ is closed under concatenations. Since $\mathcal{V}_\alpha$ is a generator for (T_1, f_α), $\lim_{n\to\infty} \operatorname{diam}(\mathcal{V}_n) = 0$, so $\mathcal{V}_n$ is a separating sequence of interval covers. $\square$

The notations in the following Lemma are those introduced in Section 4.

Lemma 1 *Let $a \in I\!N_+^{I\!N}$ be an infinite sequence of positive integers and let $u \in \mathbf{2}^*$, $|u| \geq a_0$. Let $n \geq 0$ be the unique integer with $|W_n| < |u| \leq |W_{n+1}|$, and let $1 \leq p \leq a_n$ be the unique integer with $p|W_n| < |u| \leq (p+1)|W_n|$. Then $u \in \mathcal{L}(S_a)$ iff either*

$$p = a_n \ \ \textit{and} \ \ (u \sqsubseteq W_{n+1}W_{n+1} \quad \textit{or} \quad u \sqsubseteq W_{n+1}W_nW_{n+1}),$$

or

$$p < a_n \ \ \textit{and} \ \ u \sqsubseteq W_n^k W_{n-1} W_n^l \quad \textit{for some} \ \ k, l \geq 0, \ \ k + l = p + 2.$$

Proof: Suppose first $p = a_n$. Then u is a subword of some $v \in \{W_n, W_{n+1}\}^*$ such that every occurrence of W_n in v is immediately preceded by an occurrence of W_{n+1} or some suffix, and also immediately succeeded by an occurrence of W_{n+1} or some prefix. Thus if $u \in \mathcal{L}(S_a)$, then either $u \sqsubseteq W_{n+1}W_{n+1}$ or $u \sqsubseteq W_{n+1}W_nW_{n+1}$, since $|u| \leq |W_{n+1}|$. On the other hand $W_{n+1}W_{n+1} \sqsubseteq W_{n+3}W_{n+1} \sqsubseteq W_{n+4}$, and $W_{n+1}W_nW_{n+1} \sqsubseteq W_{n+2}W_{n+1} \sqsubseteq W_{n+3}$, so both $W_{n+1}W_{n+1}$ and $W_{n+1}W_nW_{n+1}$ belong to $\mathcal{L}(S_a)$. Now if $p < a_n$, u is a subword of some $v \in \{W_{n-1}, W_n\}^*$ such that every occurrence of W_{n-1} is immediately preceded by $W_n^{a_n}$ or some suffix, and immediately followed by $W_n^{a_n}$ or some prefix. Thus if $u \in \mathcal{L}(S_a)$ then either $u \sqsubseteq W_n^k W_{n-1} W_n^l$ for some $k, l \geq 1$ with $k + l \leq p + 2$, or $u \sqsubseteq W_{n-1} W_n^{p+1} W_{n-1} \sqsubseteq W_{n-1} W_n^{p+2}$. On the other hand one has $W_n^{a_n} W_{n-1} W_n^{a_n+1} W_{n-1} \sqsubseteq (W_n^{a_n} W_{n-1})^{a_{n+1}} W_n^{a_n+1} W_{n-1} = W_{n+2} W_{n+1}$ $\sqsubseteq W_{n+3}$, so for every $k, l \geq 0$ with $k + l = p + 2$ one has $W_n^k W_{n-1} W_n^l \in \mathcal{L}(S_a)$. $\square$

Thus to decide whether u belongs to $\mathcal{L}(S_a)$, it suffices to verify whether it occurs in a test word constructed from the sequence (W_i). Observe that in all cases the test word has length at most $5|u|$: indeed if $p < a_n$ then $|W_n^k W_{n-1} W_n^l| \leq (p+3)|W_n| \leq 4p|W_n| \leq 4|u|$; if $p = a_n$ then $|W_{n+1} W_n W_{n+1}| \leq (2p+3)|W_n| \leq 5p|W_n| \leq 5|u|$.

We now encode a sequence $a \in \mathbb{N}_+^{\mathbb{N}}$ into an infinite sequence $U(a) = 1^{a_0}01^{a_1}01^{a_2}0 \cdots \in \mathbf{2}^{\mathbb{N}}$. We say that a is of class $\mathfrak{L}$ if the language $\{U(a)_{|i}; i \geq 0\}$ belongs to $\mathfrak{L}$. Put also $U_n = 1^{a_0}0 \cdots 1^{a_{n-1}}$.

Proposition 5 *Let $a \in \mathbb{N}_+^{\mathbb{N}}$ be an infinite sequence of positive integers. If a is of class* **NLIN** *or* **REC** *so is S_a.*

Proof: Suppose one already knows an algorithm recognizing the language of initial segments of $U(a)$. We describe a 6-tape Turing machine recognizing $\mathcal{L}(S_a)$ in linear additional time. Let $u \in \{0,1\}^*$ be an input word placed on the input tape 0. Let n, p be the integers associated to $|u|$ in Lemma 1. Using the Turing machine recognizing $U(a)$, we construct the initial subword of length $2|u|$ of $U(a)$ on Tape 1 (i.e., we choose a word nondeterministically and verify that it is an initial subword of $U(a)$). This word has $U_n = 1^{a_0}0 \cdots 1^{a_{n-1}}$ as a subword, and if $p = a_n$ it even contains the subword U_{n+1}. Tapes 2, 3 and 4 are used for the construction of the sequence W_k. Suppose that at the end of the $(k-1)^{st}$ step Tape 2 contains W_{k-2} and Tape 3 contains W_{k-1}. In the k^{th} step W_k is constructed on Tape 4 using the information on Tapes 1, 2 and 3. Then in the $(k+1)^{st}$ step one constructs W_{k+1} on Tape 2, as W_{k-2} is no longer necessary. The process ends when the length of the word constructed attains the length of u. At this moment, the integers n and p have been evaluated and it is also known whether $p = a_n$ or not. All this information is stored on Tape 5. Then using Lemma 1, the appropriate test word is generated on Tape 6. If $p < a_n$, this means that integers k, and l with $k + l = p + 2$ are chosen nondeterministically, and $W_n^k W_{n-1} W_n^l$ is constructed on Tape 6. If $p = a_n$, one of the words $W_{n+1} W_{n+1}$, $W_{n+1} W_n W_{n+1}$ is constructed. Finally a pointer j to Tape 6 is chosen nondeterministically, and one checks whether the test word contains the given word u from position j onward. The most time-consuming task in this algorithm is the construction of W_n. The construction of W_k from W_{k-1} and W_{k-2} takes $2|W_k|$ operations (including the return to the beginning of the tape), so the construction of W_n takes $2(|W_1| + \cdots + |W_n|)$ operations. Since $|W_k|$ grows at least as fast as the Fibonacci sequence, we get

$2(|W_1|+\cdots+|W_n|) \leq 6|W_n|$. To construct the test word on Tape 6 takes at most $(2p+3)|W_n|$ operations and checking whether u is a subword takes $(2p+3)|W_n|$ additional operations too. The algorithm requires at most $(4p+12)|W_n| \leq 16|u|$ computational operations. Now if a is recognized in linear time the combination of the two algorithms recognizes $\mathcal{L}(S_a)$ in linear time; if a is recognized by a Turing machine the combination is a Turing machine. □

We show now that a large class of irrational rotations (including all quadratic irrationals) is of class **NLIN**. An integer sequence a_n is called *polynomial* if there exists a polynomial $p(x)$ with integer coefficients and an integer q such that $a_n = p(n)$ for all $n > q$. An integer sequence $a \in \mathbb{N}_+^{\mathbb{N}}$ is called a *polynomial mixture* if there exists an integer m such that for every $0 \leq j < m$ the sequence $b_i = a_{im+j}$ is polynomial. An irrational α is a *Hurwitz number* if its continued fraction expansion is a polynomial mixture (see Perron [16], pp.126–131). Every root of a quadratic equation with integer coefficients is a Hurwitz number. Another remarkable Hurwitz number is $e = 2 + \langle 1,2,1,1,4,1,1,6\ldots\rangle$; in general if a,b,c,d,n are integers with $ad - bc \neq 0$ and $n \geq 1$, then $(ae^{2/n} + b)/(ce^{2/n} + d)$ is a Hurwitz number [9].

Proposition 6 *If $a \in \mathbb{N}_+^{\mathbb{N}}$ is a polynomial mixture, then a is of class* **NLIN**.

Proof: If a is a positive polynomial sequence of degree d, then $b_i = a_{i+1} - a_i$ is a polynomial sequence of degree $d-1$, and $b_i > 0$ for sufficiently large i. A polynomial sequence of degree d can be constructed in d steps. Given an input word u, the algorithm first constructs for the appropriate b_0 the word $1^{b_0}01^{b_0}01^{b_0}\cdots$ of length $|u|$ (degree 0). Then inductively, given a word $1^{c_0}01^{c_1}01^{c_2}..$, where c_i is a polynomial sequence of degree $k-1 < d$, the algorithm constructs for the appropriate b_k the word $1^{b_k}01^{b_k+c_0}01^{b_k+c_0+c_1}\cdots$ of length $|u|$. Finally, the algorithm rewrites the initial segment of fixed length q of the sequence where the polynomial rule does not yet apply. One step in this construction (with the return to the beginning of the tape) takes $4|u|$ operations. Thus to construct a polynomial sequence of degree d up to the coordinate $|u|$ takes $(4d+4)|u|$ operations. To construct a polynomial mixture with degrees $d_0, \ldots d_{m-1}$ takes $(4d_0 + \cdots 4d_{m-1} + 4m + 1)|u| + q$ operations. Checking whether the constructed word equals u takes $2|u|$ additional operations. □

Corollary 1 *Let α be a Hurwitz irrational. Then the corresponding rotation of the circle (T_1, f_α) is of class* **NLIN**.

The Hurwitz irrationals by no means exhaust the rotations of class **NLIN**. For example, if $a \in \{1,2\}^{\mathbb{N}}$, and if a itself is a Sturmian sequence whose continued fraction expansion is a polynomial mixture, then S_a is of class **NLIN** too. Another class of real numbers with this property has been studied by Shallit [17]. They are sums of the form $\sum_{k=0}^{\infty} q^{-2^k}$ where $q \geq 2$ is an integer. When $q \geq 3$ their continued fraction expansions are $B(q,\infty) = \lim_{r\to\infty} B(q,r)$ where

$B(q,1) = \langle q-1, q+1\rangle$ and if $B(q,r) = a_0 + \langle a_1,\ldots,a_n\rangle$ then $B(q,r+1) = a_0 + \langle a_1,\ldots,a_{n-1},a_n+1,a_n-1,a_{n-1},\ldots,a_1\rangle$; thus

$$B(3,\infty) = \langle 2,5,3,3,1,3,5,3,1,5,3,1,\ldots\rangle.$$

Clearly the word $1^{a_1}0\cdots 01^{a_{n-1}}01^{a_n+1}01^{a_n-1}1^{a_{n-1}}0\cdots 01^{a_1}$ can be constructed from $1^{a_1}0\cdots 01^{a_{n-1}}01^{a_n}$ in linear time, so $a$is of class **NLIN** whenever $a = B(q,\infty)$, and (T_1, f_α) is of class **NLIN** whenever $\alpha = \sum_{k=0}^{\infty} q^{-2^k}$ for some integer $q \geq 3$. When $q = 2$ one slight technical difficulty arises but the result is the same.

However, most irrational rotations are not of class **NLIN**. An obvious reason is that **NLIN** is a countable set of languages. Here are more specific results in this direction.

Proposition 7 *If S_a is of class* **NLIN** *then a is of class* **NTIME**(n^n).

Proof: By a theorem of [5] a language is nondeterministic linear iff it can be recognized by a nondeterministic multitape Turing machine in time equal to the length of the input word. If a word u is an initial subword of $U(a)$, then it has the form $u = 1^{a_0}0\cdots 01^{a_{n-1}}01^q$, where the a_i are positive integers, and $q \geq 0$. Checking that u is of this form is the first step of the algorithm. If $q = 0$, the word W_n is constructed on auxiliary tapes in linear time as in the proof of Proposition 5. If $q > 0$, we construct W_n^q instead. Then we verify whether the constructed word belongs to L_a: by assumption the amount of time is a linear function of the length of the word. If $q = 0$ then $|W_n| \leq (a_0+1)\cdots(a_{n-1}+1)$ $\leq |u|^n \leq |u|^{|u|}$. If $q > 0$ then $|W_n^q| \leq q(a_0+1)\cdots(a_{n-1}+1) \leq |u|^{n+1} \leq |u|^{|u|}$. The time needed is linear in $|u|^{|u|}$ in both cases. □

Corollary 2 *S_a is of class* **REC** *iff a is of class* **REC**.

Proof: This is a joint consequence of Proposition 5 and of the proof of the last result. □

As a consequence there exist irrational rotations that are of class **REC** but not of class **NLIN**, and some that are not of class **REC**; this fact relies solely on the existence of numbers with arbitrary continued fraction expansions, and can hardly be connected with arithmetic properties of these numbers.

In [7] it is proved that, while the Grand Sturmian subshift S_∞ is not of class **CF**, the complement of the associated language is context-free. This shows that for the circle endowed with the identity map (S_∞, σ) is far too complex an extension; actually looking for symbolic representations of the identity map on the circle does not make much sense. Here we show that S_∞ is of class **NLIN**; maybe this can be obtained as a consequence of the result of [7].

Proposition 8 *The Grand Sturmian subshift S_∞ is of class* **NLIN**.

Proof: The recognition algorithm for $\mathcal{L}(S_\infty)$ repeats several steps.
Step 1: One determines whether u is one of the words 0^k10^l, 1^k01^l, $(01)^k$ or one of their subwords. If so, the algorithm ends with the answer yes, otherwise it continues with Step 2.
Step 2: One determines whether there exists $q > 0$ such that for every occurrence of 10^k1 in u either $k = q$ or $k = q+1$, and neither the initial nor the final segment of u contains more than q zeroes. If the condition is satisfied, the algorithm continues with Step 3. If not, the same test is performed with zeroes and ones interchanged. If both tests fail, the algorithm ends with the answer no.
Step 3: Every word 0^q1 is replaced by 0, and the remaining $0'$s are replaced by 1. However, the procedure handles differently the beginning and end of u. If u begins with 0^k1, then this word is replaced by 0, if u ends with 0^k then this word is replaced nondeterministically either by 0 (if $k \leq q$) or by 10 (if $k \geq 2$). The algorithm then repeats the procedure starting in Step 1 with the word $\nu(u)$ obtained by these replacements.
It is clear that Steps 1, 2, and 3 take linear time $c|u|$. For the length of the word $\nu(u)$ we have $|u| = (q+1)|\nu(u)|_0 + |\nu(u)|_1$. Since $q > 0$, we get $|\nu(u)| \leq \frac{5}{6}|u|$, so the whole algorithm works in time $c|u|/(1-5/6) = 6c|u|$. $\Box$

Finally here is a number-theoretic remark. Among the Sturmian subshifts of class **NLIN** some correspond to algebraic numbers (the quadratic numbers) and some to transcendental numbers (e, Shallit's examples). On the other hand we do not know anything about the numbers generating Sturmian subshifts that are not of class **NLIN**. For instance does this last family contain algebraic numbers? It is possible to prove that Sturmian subshifts associated to irrational algebraic numbers are of class **REC**.

Rather naturally this open question corresponds to a gap in the theory of continued fraction expansions.

References

[1] R.Adler, B.Weiss: Similarity of automorphisms of the torus, Memoir Amer. Math. Soc. **98** (1970).

[2] P.Arnoux: Représentation géométrique des systèmes dynamiques symboliques. Habilitation, Université d'Aix-Marseille II, 1991.

[3] F.Blanchard: β-expansions and symbolic dynamics. *Theoret. Comput. Sci.* **65** (1989) 131–141.

[4] F.Blanchard, G.Hansel: Languages and subshifts, in: *Automata on Infinite Words*, M.Nivat and D.Perrin eds., Lecture Notes in Computer Science 192, 138–146, Springer-Verlag, Berlin, 1985.

[5] R.V.Book, S.A.Greibach: Quasi-realtime languages, *Math. Systems Th.* **4** (1970) 97–111.

[6] R.L.Devaney: *An Introduction to Chaotic Dynamical Systems.* Addison-Wesley, 1989.

[7] S.Duluc, D.Gouyou-Beauchamp: Sur les facteurs des suites de Sturm, *Theoret. Comput. Sci.* **71** (1990), 381–400.

[8] J.E.Hopcroft, J.D.Ullman: *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley, 1979.

[9] A. Hurwitz: Über die Kettenbrüche, deren Teilnenner arithmetische Reihen bilden, *Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich* **41** (1896), Jubelband II, 34–64; *Werke*, Vol. 2, 276–302.

[10] P.Kůrka: Languages, equicontinuity and attractors in cellular automata, to appear in *Ergodic Th. Dyn. Sys.* 1996

[11] P.Kůrka: Zero-dimensional dynamical systems, formal languages and universality, preprint, 1996

[12] D.Lind, B.Marcus: *An Introduction to Symbolic Dynamics and Coding.* Cambridge University Press. Cambridge, 1995.

[13] J.Milnor, W.Thurston: On iterated maps of the interval, in: *Dynamical Systems*, J.C.Alexander ed., Lecture Notes in Math. 1342, 465–563, Springer, Berlin (1988).

[14] M.Morse, G.A.Hedlund: Symbolic dynamics II: Sturmian trajectories, *Amer. J. Math.* **62** (1940) 1–42.

[15] Ch.H.Papadimitriou: *Computational Complexity.* Addison-Wesley, Reading 1994.

[16] O.Perron: *Die Lehre von den Kettenbrüchen.* B.G.Teubner, Leipzig 1929.

[17] J.Shallit: Simple continued fractions for some irrational numbers. *J. Number Theory* **11** (1979) 209–217.

[18] K.Wagner, G.Wechsung: *Computational Complexity.* VEB Deutscher Verlag der Wissenschaften, Berlin 1986.

F. Blanchard: Institut de Mathématiques de Luminy-CNRS
Case 930, 163 avenue de Luminy, F-13288 Marseille cedex 09, France
P. Kůrka: Institut de Mathématiques de Luminy- CNRS
permanent address: Faculty of Mathematics and Physics, Charles University in Prague, Malostranské nám. 25, CZ-118 00 Praha 1, Czechia.